Bayesian modelling of clusters of galaxies from multi-frequency pointed Sunyaev–Zel'dovich observations

Farhan Feroz*, Michael P. Hobson, Jonathan T. L. Zwart,
Richard D. E. Saunders and Keith J. B. Grainge
Astrophysics Group, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge, CB3 0HE

ABSTRACT

We present a Bayesian approach to modelling galaxy clusters using multi-frequency pointed observations from telescopes that exploit the Sunyaev–Zel'dovich effect. We use the recently developed MULTINEST technique (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2008) to explore the high-dimensional parameter spaces and also to calculate the Bayesian evidence. This permits robust parameter estimation as well as model comparison. Tests on simulated Arcminute Microkelvin Imager observations of a cluster, in the presence of primary CMB signal, radio point sources (detected as well as an unresolved background) and receiver noise, show that our algorithm is able to analyse jointly the data from six frequency channels, sample the posterior space of the model and calculate the Bayesian evidence very efficiently on a single processor. We also illustrate the robustness of our detection process by applying it to a field with radio sources and primordial CMB but no cluster, and show that indeed no cluster is identified. The extension of our methodology to the detection and modelling of multiple clusters in multi-frequency SZ survey data will be described in a future work.

Key words: methods: data analysis – methods: statistical – cosmology:observations – galaxies: clusters: general – cosmic microwave background

1 INTRODUCTION

Clusters of galaxies are the most massive gravitationally bound objects in the Universe and as such are critical tracers of the formation of large-scale structure. The size and formation history of massive clusters is such that the ratio of cluster gas mass to total mass is expected to be representative of the universal ratio Ω_b/Ω_m, once the relatively small amount of baryonic matter in the cluster galaxies is taken into account (see e.g. White et al. 1993). The comoving number density of clusters as a function of mass and redshift is expected to be particularly sensitive to the cosmological parameters σ_8 and Ω_m (see e.g. Battye & Weller 2003). This has been predicted both analytically (see e.g. Press & Schechter 1974; Sheth et al. 2001) and from large-scale numerical simulations (see e.g. Jenkins et al. 2001; Evrard et al. 2002), but cluster number densities have not yet been measured at redshifts $z \gtrsim 1$, because of the basic problem of the dimming of surface brightness with redshift. Moreover, optical observations are confused by foreground galaxies, and both optical and X-ray observations are biased towards strong mass concentrations such as clumps.

The Sunyaev–Zel'dovich (SZ, Sunyaev & Zeldovich 1970, 1972; see e.g. Birkinshaw 1999 and Carlstrom, Holder & Reese 2002 for reviews) effect produces secondary anisotropies in the cosmic microwave background (CMB) radiation through inverse-Compton scattering from the electrons in the hot intracluster gas (which radiates via thermal Bremsstrahlung in the X-ray waveband), and the subsequent transfer of some of the energy of the electrons to the low-energy photons. Pointed SZ observations of clusters have been routine for some time (see e.g. Birkinshaw, Gull & Moffet 1981; Birkinshaw, Gull & Hardebeck 1984; Uson 1986, Jones et al. 1993), Grainge et al. 1996, Grainge et al. 2002a,b, Cotter et al. 2002a,b, Grainger et al. 2002, Saunders et al. 2003, Jones et al. 2005; Carlstrom, Joy & Grego 1996, Grego et al. 2000, Patel et al. 2000, Reese et al. 2000, Joy et al. 2001, Reese et al. 2002, Bonamente et al. 2006, LaRoque et al. 2006). By fitting simple parametric cluster models to the observed data set, one would like to make model-dependent inferences about the cluster parameters, i.e. to calculate the probability distribution of these parameters. We also wish to compare different cluster models to enhance our astrophysical understanding. These tasks are most conveniently carried out through Bayesian inference.

Marshall et al. (2003) presented a Bayesian approach for the joint analysis of pointed SZ and weak lensing data. They used a highly effective but computationally intensive Markov Chain Monte Carlo sampler to explore the high-dimensional parameter space and employed the thermodynamic integration technique to calculate the Bayesian evidence. We have since ex-

* E-mail: f.feroz@mrao.cam.ac.uk

tended this approach to incorporate the highly efficient parameter-space sampling method, MULTINEST (Feroz & Hobson 2008; Feroz et al. 2008), allowing us to analyse efficiently cluster observations in the presence of multiple radio sources. The analysis can now be done on multi-frequency SZ data. The new generation of SZ instruments including ACT (Kosowsky 2006), AMI (AMI Consortium: Zwart et al. 2008), AMiBA (Li et al. 2006), APEX-SZ (Dobbs et al. 2006), CARMA (www.mmarray.org) and SPT (Ruhl et al. 2004), all have multiple frequency channels. Our algorithm includes a more sophisticated model for the radio sources by allowing for their spectral indices to be non-flat. We also take into account the noise contribution coming from the population of faint unsubtracted radio sources. Furthermore, the Bayesian evidence is now handled comprehensively and used for objective and quantitative detection of clusters (following Hobson & McLachlan 2003) as well as for model selection between different cluster models. In this paper we restrict our focus to the analysis of pointed SZ observations, but our basic methodology can be extended to the detection and modelling of multiple clusters in multi-frequency SZ survey data, and will be described in a forthcoming publication.

In Section 2 we briefly describe the AMI telescope. In Section 3 we give an introduction to the Bayesian inference. Sections 4 and 5 describe the SZ effect and our analysis methodology including those features described above. In Section 6 we apply our algorithm to simulated SZ cluster data and we present our conclusions in Section 7.

2   THE ARCMINUTE MICROKELVIN IMAGER

Although the cluster modelling method we present in this paper is quite general in nature, a principal goal of this work has been to develop an efficient and robust technique for analysing cluster SZ observations made by the Arcminute Microkelvin Imager (AMI, Kneissl et al. 2001, AMI Consortium: Zwart et al. 2008). This instrument consists of a pair of interferometer arrays operating currently with six frequency channels spanning 13.9–18.2 GHz for observations on angular scales of $30''$–$10'$. The telescope is aimed principally at SZ imaging of clusters of galaxies (AMI Consortium: Grainge et al. 2006). In order to couple to the extended SZ flux from cluster gas structures, the Small Array (SA) has a large filling factor and thus excellent temperature sensitivity. The dominant contaminant in SZ observations at these Rayleigh-Jeans frequencies are the radio sources (see also Section 5.3). The Large Array (LA) deliberately has higher resolution and better flux sensitivity than the SA and, observing simultaneously with the SA both in time and frequency, allows subtraction of such radio point sources. SA and LA parameters are summarized in Table 1.

3   BAYESIAN INFERENCE

Our cluster modelling methodology is built upon the principles of Bayesian inference; we now give a summary of this framework. Bayesian inference methods provide a consistent approach to the estimation of a set of parameters Θ in a model (or hypothesis) H for the data D. Bayes' theorem states that

$$\Pr(\Theta|D, H) = \frac{\Pr(D|\Theta, H)\Pr(\Theta|H)}{\Pr(D|H)}, \qquad (1)$$

Table 1. AMI technical summary.

	SA	LA
Antenna diameter	3.7 m	12.8 m
Number of antennas	10	8
Baseline lengths (current)	5–20 m	18–110 m
Primary beam (15.7 GHz)	$20'.1$	$5'.5$
Synthesized beam	$\approx 3'$	$\approx 30''$
Flux sensitivity	30 mJy s$^{-1/2}$	3 mJy s$^{-1/2}$
Observing frequency	13.9–18.2 GHz	
Bandwidth	4.3 GHz	
Number of channels	6	
Channel bandwidth	0.72 GHz	

where $\Pr(\Theta|D, H) \equiv P(\Theta)$ is the posterior probability distribution of the parameters, $\Pr(D|\Theta, H) \equiv \mathcal{L}(\Theta)$ is the likelihood, $\Pr(\Theta|H) \equiv \pi(\Theta)$ is the prior probability distribution, and $\Pr(D|H) \equiv \mathcal{Z}$ is the Bayesian evidence.

In parameter estimation, the normalising evidence factor is usually ignored, since it is independent of the parameters Θ, and inferences are obtained by taking samples from the (unnormalised) posterior using standard MCMC sampling methods, where at equilibrium the chain contains a set of samples from the parameter space distributed according to the posterior. This posterior constitutes the complete Bayesian inference of the parameter values, and can, for example, be marginalised over each parameter to obtain individual parameter constraints.

In contrast to parameter estimation, for model selection the evidence takes the central role and is simply the factor required to normalize the posterior over Θ:

$$\mathcal{Z} = \int \mathcal{L}(\Theta)\pi(\Theta)d^D\Theta, \qquad (2)$$

where D is the dimensionality of the parameter space. As the average of the likelihood over the prior, the evidence is larger for a model if more of its parameter space is likely and smaller for a model with large areas in its parameter space having low likelihood values, even if the likelihood function is very highly peaked. Thus, the evidence automatically implements Occam's razor: a simpler theory with compact parameter space will have a larger evidence than a more complicated one, unless the latter is significantly better at explaining the data. The question of model selection between two models H_0 and H_1 can then be decided by comparing their respective posterior probabilities, given the observed data set D, via the model selection ratio R:

$$R = \frac{\Pr(H_1|D)}{\Pr(H_0|D)} = \frac{\Pr(D|H_1)\Pr(H_1)}{\Pr(D|H_0)\Pr(H_0)} = \frac{\mathcal{Z}_1}{\mathcal{Z}_0}\frac{\Pr(H_1)}{\Pr(H_0)}, \qquad (3)$$

where $\Pr(H_1)/\Pr(H_0)$ is the *a priori* probability ratio for the two models, which can often be set to unity but occasionally requires further consideration.

Various alternative information criteria for astrophysical model selection are discussed by Liddle (2007), but the evidence remains the preferred method. However, evaluation of the multidimensional integral in (2) is a challenging numerical task. Standard techniques like thermodynamic integration are extremely computationally expensive, which makes evidence evaluation at least an order-of-magnitude more costly than parameter estimation. Some fast approximate methods have been used for evidence evaluation, such as treating the posterior as a multivariate Gaussian centred at its peak (see e.g. Hobson et al. 2002), but this approximation is clearly a poor one for multimodal posteriors (except perhaps if one

performs a separate Gaussian approximation at each mode). The Savage-Dickey density ratio has also been proposed (see e.g. Trotta 2007) as an exact, and potentially faster, means of evaluating evidences, but is restricted to the special case of nested hypotheses and a separable prior on the model parameters.

The nested sampling approach, introduced by Skilling (2004), is a Monte-Carlo method targeted at the efficient calculation of the evidence, but also produces posterior inferences as a by-product. Feroz & Hobson (2008) and Feroz et al. (2008) built on this nested sampling framework, and have recently introduced the MULTI-NEST algorithm which is very efficient in sampling from posteriors that may contain multiple modes and/or large (curving) degeneracies, and also calculates the evidence. This technique has greatly reduced the computational cost of Bayesian parameter estimation and model selection, and is employed in this paper.

4 SUNYAEV–ZEL'DOVICH EFFECT FROM CLUSTERS

The primary anisotropies in the CMB are roughly one part in 10^5 and reflect the intrinsic non-uniformity of the matter and radiation just before the Universe cooled sufficiently to form transparent atomic gas. The secondary anisotropies in the CMB are due to processes affecting the CMB after its emission. Much the most relevant for this work is the thermal SZ effect.

The gas temperature in galaxy clusters is 10^7–10^8 K. With the radius r of the cluster, the electron number density n_e, the optical depth for Thomson scattering through the centre of the cluster is $\tau \approx 2 r n_e \sigma_T \sim 10^{-2}$, where σ_T the Thomson scattering cross-section. On average, energy is transferred from the electrons to the scattered CMB photons, with the fractional energy increase approximately equal to $k_B T/m_e c^2 \sim 10^{-2}$ with k_B the Boltzmann constant, T and m_e the electron mass and temperature respectively and c the speed of light. These factors combine to give fractional CMB temperature fluctuations of order 10^{-4}. A full treatment (see e.g. Birkinshaw 1999) yields the following modification to the CMB surface brightness in the direction of the reservoir of electrons:

$$\delta I_\nu = f(\nu) y B_\nu(T_{CMB}), \qquad (4)$$

where $B_\nu(T_{CMB})$ is the blackbody spectrum at $T_{CMB} = 2.726$K, while the frequency-dependent function $f(\nu)$ in the limit of non-relativistic plasma is given as

$$f(\nu) = \frac{x \left[x \coth(x/2) - 4 \right]}{1 - e^{-x}}, \qquad (5)$$

where

$$x = \frac{h_p \nu}{k_B T_{CMB}}, \qquad (6)$$

with h_p the Planck constant. At frequencies below 217 GHz, the SZ effect is observed as a flux decrement. The 'Comptonisation' parameter y is the integral of the gas pressure along the line of sight l through the cluster:

$$y = \frac{\sigma_T}{m_e c^2} \int n_e k_B T \, dl. \qquad (7)$$

The integral $\mathcal{Y}$ of this Comptonisation parameter over a cluster's solid angle $d\Omega = dA/D_\theta^2$, with D_θ the angular-diameter distance to the cluster, is proportional to the total thermal energy content of the cluster:

$$\mathcal{Y} = \int y \, d\Omega = \frac{\sigma_T}{m_e c^2} \int n_e k_B T \, dl \, d\Omega \propto \frac{1}{D_\theta^2} \int n_e T \, dV, \quad (8)$$

where $dV = dA \, dl$ is an element of comoving volume. The integrated SZ surface brightness simply depends on the total cluster mass M, since,

$$\mathcal{Y} \propto \frac{\langle T \rangle}{D_\theta^2} \int n_e \, dV \propto \frac{M^{2/3}}{D_\theta^2} M \propto \frac{M^{5/3}}{D_\theta^2}, \qquad (9)$$

where we have assumed that the cluster gas mass is proportional to its total mass, i.e. $M_g = f_g M$.

One final property of the SZ effect is particularly significant: for a given cluster, the SZ surface brightness is *independent of redshift*. The redshift only enters via the angular-diameter distance $D_\theta(z)$, which at intermediate redshifts ($0.5 \lesssim z \lesssim 6$) is only weakly dependent on redshift for a concordance cosmology. Therefore, an SZ survey is expected to find all high-redshift clusters above some mass threshold with little dependence on the redshift.

5 MODELLING INTERFEROMETRIC SZ DATA

The majority of SZ observations to date have been made with interferometers (see section 1). These instruments have a number of advantages over single-dish telescopes (AMI Consortium: Zwart et al. 2008, and references therein), including their relative insensitivity to atmospheric emission (Lay & Halverson 2000), lack of required receiver stability, and the ease with which systematic errors such as ground spill (Watson et al. 2003) and radio-source contamination (see e.g. Grainger et al. 2002) can be minimised.

Assuming a small field size, an interferometer operating at a single frequency ν measures samples from the complex visibility plane $\widetilde{I}_\nu(\boldsymbol{u})$. This is given by the weighted Fourier transform of the surface brightness $I_\nu(\boldsymbol{x})$,

$$\widetilde{I}_\nu(\boldsymbol{u}) = \int A_\nu(\boldsymbol{x}) I_\nu(\boldsymbol{x}) \exp(2\pi i \boldsymbol{u} \cdot \boldsymbol{x}) \, \mathrm{d}^2\boldsymbol{x}, \qquad (10)$$

where $\boldsymbol{x}$ is the position relative to the phase centre, $A_\nu(\boldsymbol{x})$ is the (power) primary beam of the antennas at the observing frequency ν (normalised to unity at its peak), and $\boldsymbol{u}$ is a baseline vector in units of wavelength. Interferometers effectively measure spatial structures in the surface brightness δI_ν as the large-angular-scale 'DC' level at the centre of the uv-plane is never sampled. The positions in the uv-plane at which the function $\widetilde{I}_\nu(\boldsymbol{u})$ is sampled are determined by the physical positions of the antennas and the direction of the field on the sky. The samples $\boldsymbol{u}_i$ lie on a series of curves, called uv-tracks.

In our model, the we assume the measured interferometer visibilities contain contributions from the cluster SZ signal, radio point sources, primordial CMB anisotropies and instrumental noise; these contributions are discussed below. In short, however, each interferometer visibility is considered to consist of a signal and generalised noise:

$$V_\nu(\boldsymbol{u}_i) = \widetilde{I}_\nu(\boldsymbol{u}_i) + N_\nu(\boldsymbol{u}_i), \qquad (11)$$

where the signal part contains the contributions from the SZ cluster and identified radio point sources, and the generalised noise part contains the contributions from the unresolved background of radio point sources, primordial CMB anisotropies and instrumental noise.

5.1   Cluster model

To determine the contribution to the visibility data from the cluster SZ signal, one needs to calculate the Comptonisation parameter of the cluster given in (7), for which one must assume a cluster geometry as well as temperature and pressure profiles for the cluster gas.

For the cluster geometry, spherically-symmetric models are a reasonable first approximation. Ellipticity can be added easily through a coordinate transformation (see Marshall 2003), but we will not pursue that here.

The simplest gas temperature model assumes a single temperature which is in good agreement with low-resolution X-ray emission data (see e.g. Sarazin 1988). One could alternatively assume a polytropic temperature with

$$p_g \propto \rho_g^\gamma \Rightarrow T \propto \rho_g^{\gamma-1}, \tag{12}$$

with γ being the polytropic index, and p_g and ρ_g the gas pressure and density respectively. The polytropic model has been found to provide a good fit to simulated clusters (see e.g. Komatsu & Seljak 2001). However, in this work we take the cluster gas to be isothermal.

The cluster gas density is often modelled with a β-model (Cavaliere & Fusco-Femiano 1976, 1978). This has the density profile

$$\rho_g(r) = \frac{\rho_g(0)}{\left[1 + (r/r_c)^2\right]^{\frac{3\beta}{2}}}, \tag{13}$$

where r_c is the core radius at which the profile turns over into a region of approximately constant density, while the outer logarithmic slope of the profile is 3β.

For simplicity, we will also assume here that the cluster gas is in hydrostatic pressure equilibrium with the total gravitational potential Φ of the cluster, which will be dominated by the dark matter. Assuming spherical symmetry, the gravitational potential Φ must thus satisfy

$$\frac{d\Phi}{dr} = -\frac{1}{\rho_g}\frac{dp}{dr}. \tag{14}$$

Assuming the cluster gas to be an ideal gas with temperature T, and its distribution to be spherically symmetric, (14) becomes

$$\frac{d\log\rho_g}{d\log r} = -\frac{GM(r)\mu}{k_B T r}, \tag{15}$$

where μ is the mass per particle, approximately 0.6 times the proton mass (see Marshall et al. 2003), G is the universal gravitational constant and $M(r)$ is the total mass internal to radius r. We note that an alternative approach would be to assume a dark matter profile, such as the NFW profile (Navarro et al. 1997), and use the assumption of hydrostatic equilibrium to determine the corresponding gas density profile (Marshall et al. 2003), but we will not pursue that here.

Taking together the assumptions of hydrostatic equilibrium, spherical symmetry, a beta profile for the gas density and an isothermal temperature profile, (13) and (15) lead to

$$M(r) = \frac{3\beta r^3}{r_c^2 + r^2}\frac{k_B T}{\mu G}. \tag{16}$$

This can be used to calculate the total cluster mass out to r_{200}, the radius inside which the average total density is 200 times the critical density ρ_{crit}, as

$$
\begin{aligned}
M_{200} &= \frac{4\pi}{3}r_{200}^3(200\rho_{\text{crit}}) \\
&= \frac{3\beta r_{200}^3}{r_c^2 + r_{200}^2}\frac{k_B T}{\mu G}.
\end{aligned} \tag{17}
$$

For the sake of brevity, we shall refer to M_{200} simply as the total mass of the cluster, and write it as M.

Although the beta model (13) for the gas density profile is expressed in terms of $\rho_g(0)$, the more interesting parameter is the total gas mass $M_g(r)$ inside a certain radius r. This can easily be accommodated by instead treating $M_g(r_{200}) \equiv M_g$ as a model parameter and recovering $\rho_g(0)$ by numerically integrating the gas density profile to a given radius and subsequently normalizing the gas mass within this radius. For this work, a radius of 1 h^{-1}Mpc was used as the limit of integration for the gas density profile.

Thus, the parameters of our cluster model are taken to be T, M_g, β and r_c, along with the position (x_c, y_c) of its centroid on the sky and its redshift z. For a given set of values for these parameters, a map of the predicted Comptonisation parameter can be calculated by evaluating the line-of-sight integral of the projected gas density profile:

$$y(s) = \frac{\sigma_T}{m_e c^2}\int_{-\infty}^{\infty} n_e k_B T\, dl, \tag{18}$$

where $s = \theta D_\theta$ is the projected radius, such that $r^2 = s^2 + l^2$. Since

$$\rho_g = n_e \mu_e, \tag{19}$$

where $\mu_e = 1.14 m_p$ (Mason & Myers 2000; Jones et al. 1993) is the gas mass per electron, (18) becomes

$$y(s) \propto \int_{-r_{\text{lim}}}^{r_{\text{lim}}} \rho_g(r)\, T\, dl, \tag{20}$$

where r_{lim} defines the radial limit of the integration. We set r_{lim} to $20h^{-1}$Mpc which is sufficiently large even for low values of β.

It only remains to specify the prior $\pi(\Theta_c)$ on our cluster model parameters $\Theta_c \equiv (x_c, y_c, \beta, r_c, T, M_g, z)$. Pointed SZ observations will typically be directed towards (putative) clusters already observed in the X-ray or optical bands. Ideally, one would perform a joint analysis of these data sets (see e.g. Marshall et al. 2003) to constrain cluster parameters. For our present purpose of analysing SZ data alone, however, one can consider observations in other wavebands simply to provide (joint) priors on cluster parameters in the analysis of the subsequent SZ data.

For simplicity, we will assume throughout this paper that the prior is separable, except in the parameters M_g and z, such that

$$\pi(\Theta_c) = \pi(x_c)\pi(y_c)\pi(\beta)\pi(r_c)\pi(T)\pi(M_g, z). \tag{21}$$

We use Gaussian priors on cluster position parameters, centered on the pointing centre and with standard deviation 1 arcmin. We adopt uniform priors $\pi(r_c) = \mathcal{U}(0, 1000)h^{-1}$kpc and $\pi(\beta) = \mathcal{U}(0.3, 1.5)$ on the cluster core radius and β, the outer logarithmic slope of the gas density profile. A Gaussian prior on cluster temperature can be adopted with the mean and standard deviation coming from X-ray or optical (from a velocity dispersion) measurements. In the absence of any X-ray or optical observation of the cluster, we use a uniform prior $\pi(T) = \mathcal{U}(0, 20)$ keV on the temperature. Similarly, optical observations could, in principle, provide a joint prior on the mass and redshift of the cluster. In the absence of such observations, one can instead assign the prior based upon an assumed cluster mass function, as we now discuss.

Since it requires little further work, let us consider the very general case in which the cluster gas fraction f_g is also unknown, but we have some prior $\pi(f_g)$ on its value, perhaps from previous observations (we will assume for simplicity that f_g is independent of redshift, although it is easy to relax this condition also). The gas mass of a cluster is related to its 'total mass' M ($\equiv M_{200}$) by $M_g = f_g M$ and hence it is straightforward to show that

$$\pi(M_g, z) = \int_0^1 \pi(f_g)\pi(M = M_g/f_g, z)\frac{df_g}{f_g}, \qquad (22)$$

where $\pi(M, z)$ is the joint prior on the total cluster mass and redshift. In turn, we take the latter to be equal to some assumed cluster mass function, $d^2n/dM\,dz$, appropriately normalised over some ranges $M_{min} < M \leqslant M_{max}$ and $z_{min} < z \leqslant z_{max}$. It is worth noting that, in the special case where the cluster gas fraction is known, one simply has $\pi(M_g, z) \propto \pi(M = M_g/f_g, z)$.

One possibility for $d^2n/dM\,dz$ is the Press–Schechter (Press & Schechter 1974) mass function. Numerical simulations have shown that the Press–Schechter mass function overestimates the abundance of high-mass clusters and underestimates those of low mass (Sheth et al. 2001), but overall it still provides an adequate fit to N-body simulations (Jenkins et al. 2001). In particular, a reasonable fit is obtained for the Press-Schecter mass function with $\sigma_8 = 0.8$, which is plotted in Figure 1, along with some samples drawn from it for illustration.

Another possibility is simply to assume a separable prior in M and z, namely $\pi(M, z) = \pi(M)\pi(z)$, where each factor has some simple functional form such that their product gives a reasonable approximation to the Press–Schechter mass function. We shall assume such a form in our analysis of simulated SZ data in Section 6, where $\pi(M)$ will be taken to be uniform in $\log M$ in the range $M_{min} = 10^{13}h^{-1}M_\odot$ to $M_{max} = 5 \times 10^{15}h^{-1}M_\odot$, and zero outside this range. Moreover, in Section 6, we will also assume simply that the cluster redshift and gas fraction are known, which is equivalent to imposing delta function priors on z and f_g.
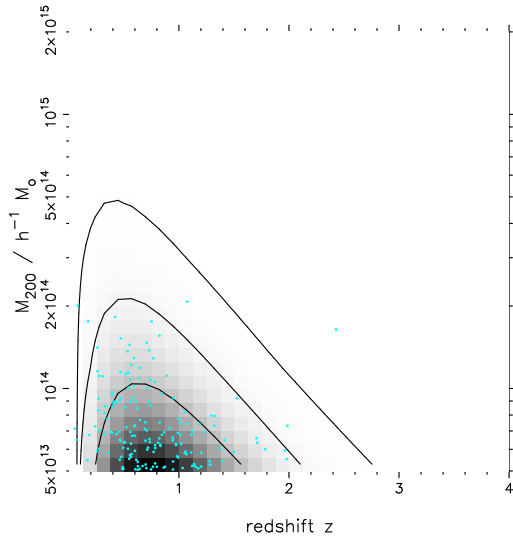
Figure 1. The Press-Schechter mass function with $\sigma_8 = 0.8$, together with some samples drawn from it for illustration. The contours enclose 68%, 95% and 99% of the probability.

5.2 Resolved radio point-sources model

A key issue for SZ cluster observations is the effect of contaminating radio sources. The most obvious problem is that the emission from source(s) coincident with the cluster centre can mask the SZ decrement. A more subtle but equally disastrous problem is that emission from a radio source lying on a negative sidelobe of the interferometer response (centered at the cluster) can mask an SZ decrement. AMI's observational approach to this problem is robust: with the RT and VSA in the past, and with AMI LA at present, we provide enough telescope sensitivity at high angular resolution to measure the flux densities at positions of the contaminating sources down to a faint limiting flux density S_{lim}. We then handle these sources as follows.

The visibility due to each of these radio sources can be calculated as:

$$\widetilde{I}_\nu^S(\boldsymbol{u}) = \int A_\nu(\boldsymbol{x})S_\nu(\boldsymbol{x})\exp(2\pi i\boldsymbol{u}\cdot\boldsymbol{x})\,d^2\boldsymbol{x}, \qquad (23)$$

where $S_\nu(\boldsymbol{x})$ is the source flux at a position $\boldsymbol{x}$ relative to the phase centre. Most of these foreground radio sources appear as 'point' sources in the data, in which case the visibility can be calculated analytically:

$$\begin{aligned}\widetilde{I}_\nu^S(\boldsymbol{u}) &= \int A_\nu(\boldsymbol{x})S_\nu\delta(\boldsymbol{x}-\boldsymbol{x}_s)\exp(2\pi i\boldsymbol{u}\cdot\boldsymbol{x})\,d^2\boldsymbol{x}\\ &= S_\nu A_\nu(\boldsymbol{x}_s)e^{i\phi}, \qquad (24)\end{aligned}$$

where $\boldsymbol{x}_s$ is the position of the radio point source and $\phi = 2\pi\boldsymbol{u}\cdot\boldsymbol{x}_s$ is the phase angle. If the radio source is extended, then a Gaussian profile can be used as a simple model for its flux distribution and (23) can again be solved analytically. For this work, however, we assume all the radio sources to be point sources.

For SZ observations made at more than one frequency, the change of source flux with frequency ν also needs to be taken into account. Here we assume a power-law dependence with spectral index α such that

$$S_\nu = S_0\left(\frac{\nu}{\nu_0}\right)^{-\alpha}, \qquad (25)$$

where ν_0 is some reference frequency (usually that of the lowest observed frequency channel).

For each identified radio point source, we impose the same priors on its parameters $\boldsymbol{\Theta}_s \equiv (x_s, y_s, S_0, \alpha)$. We assume throughout that the prior is separable:

$$\pi(\boldsymbol{\Theta}_s) = \pi(x_s)\pi(y_s)\pi(S_0)\pi(\alpha). \qquad (26)$$

For the source position parameters, we use a delta function prior centered on the measured position that is assumed known from higher-resolution observations with the AMI LA. Although, such observations also yield a source flux measurement that can be used to give a very narrow prior on S_0, for the analysis in this paper, we use a uniform prior $\mathcal{U}(0, 20)$ mJy to allow our algorithm to determine how accurately the source fluxes can be fitted.

For the spectral index α, a little more care is required in the choice of prior. Although radio source spectra typically fall with an index of 0.7, they differ widely because of synchrotron self-absorption and ageing effects, and so can vary between ≈ 2 (steeply falling), through to ≈ 0 (flat) and even ≈ -1 (rising). Flat- and rising-spectrum sources tend to be more variable in time, and we can straightforwardly account for this by increasing a source's flux uncertainty for flat or rising sources. Waldram et al. (2007) found 15–22 GHz spectral indices for a sample of 110 9C (Waldram et al. 2003) sources. The modal spectral index was found

to be 0.5. We use the distribution (Figure 2) of these spectral indices as the prior $\pi(\alpha)$, binned onto a grid with $\Delta\alpha = 0.1$. We expect a source's spectral index to be well-constrained by any available long-baseline data, such as that from the LA, and so inform the prior.

Finally, since numerous radio sources are typically identified in each field, we combine the parameters $\boldsymbol{\Theta}_s$ for each source in to a single radio-sources parameter set $\boldsymbol{\Psi} = (\boldsymbol{\Theta}_{s1}, \boldsymbol{\Theta}_{s2}, \dots)$ and we assume the priors on the parameters of different sources are separable.

5.3   Generalised noise model and likelihood function

As mentioned previously, there are three components that contribute to the generalised 'noise' on the visibilities (11): instrumental noise, primordial CMB anisotropies and confusion noise from the background of unsubtracted radio sources. These contributions are discussed below and define the likelihood function for the data.

Following Hobson & Maisinger (2002), it is convenient first to place the $N_{\mathrm{vis},\nu}$ observed complex visibilities $V_\nu(\boldsymbol{u}_i)$ in frequency channel ν into a real data vector $\boldsymbol{d}_\nu$ with components

$$\boldsymbol{d}_{\nu,\mathrm{i}} = \begin{cases} \mathrm{Re}\{V_\nu(\boldsymbol{u}_i)\} & (i \leqslant N_{\mathrm{vis},\nu}) \\[2mm] \mathrm{Im}\{V_\nu(\boldsymbol{u}_{i-N_{\mathrm{vis},\nu}})\} & (N_{\mathrm{vis},\nu} + 1 \leqslant i \leqslant 2N_{\mathrm{vis},\nu}). \end{cases} \tag{27}$$

It will also be useful to define the total data vector $\boldsymbol{d} = \{\boldsymbol{d}_\nu\}$, which is the concatenation of the individual data vectors for each frequency channel ν. Similarly, one may define the real noise vectors $\boldsymbol{n}_\nu$ containing only the noise parts $N_\nu(\boldsymbol{u}_i)$ of the complex visibilities in (11), and the concatenated noise vector $\boldsymbol{n}$.

The instrumental noise and primordial CMB anisotropies are well described by Gaussian processes. The background of unresolved radio sources is, however, a Poisson process. Nonetheless, in the limit of a large number of unresolved sources, this contribution can also be well approximated as Gaussian (see below). Thus, in this paper, we shall assume a Gaussian form of the likelihood
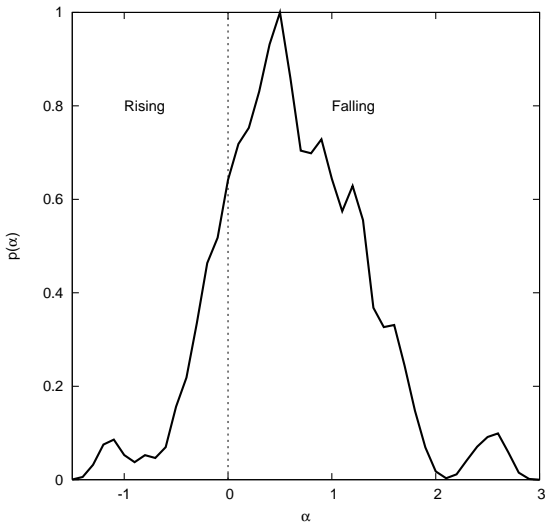
Figure 2. Spectral index distribution from Waldram et al. (2007) used as the prior on α, normalized to unit probability at maximum. The final normalisation depends on the adopted prior range.

function on the full parameter set $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_c, \boldsymbol{\Psi})$:

$$\mathcal{L}(\boldsymbol{\Theta}) = \frac{1}{Z_{\mathrm{N}}} \exp(-\tfrac{1}{2}\chi^2), \tag{28}$$

where χ^2 is the standard statistic quantifying the misfit between the observed data $\boldsymbol{d}$ and the predicted data $\boldsymbol{d}^{\mathrm{p}}(\boldsymbol{\Theta})$,

$$\chi^2 = \sum_{\nu,\nu'}(\boldsymbol{d}_\nu - \boldsymbol{d}_\nu^{\mathrm{p}})^{\mathrm{T}}(\boldsymbol{C}_{\nu,\nu'})^{-1}(\boldsymbol{d}_{\nu'} - \boldsymbol{d}_{\nu'}^{\mathrm{p}}), \tag{29}$$

in which $\boldsymbol{C}_{\nu,\nu'} \equiv \langle \boldsymbol{n}_\nu \boldsymbol{n}_{\nu'}^{\mathrm{T}} \rangle$ is the generalised noise covariance matrix relating the frequency channels ν and ν'. The normalisation factor Z_{N} in (28) is given by

$$Z_{\mathrm{N}} = (2\pi)^{(2N_{\mathrm{vis}})/2}|\boldsymbol{C}|^{1/2}, \tag{30}$$

where $\boldsymbol{C} \equiv \langle \boldsymbol{n}\boldsymbol{n}^{\mathrm{T}} \rangle$ and N_{vis} is the total number of visibilities in all the frequency channels. In this work, Z_{N} is independent of the model parameters $\boldsymbol{\Theta}$ and hence it can be safely ignored in the analysis.

In (29), the predicted data $\boldsymbol{d}_\nu^{\mathrm{p}}$ at each frequency are a function of the model parameters $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_c, \boldsymbol{\Psi})$. For a given set of parameter values, the predicted data are calculated as follows. First, the values of the cluster parameters $\boldsymbol{\Theta}_c$ are used to calculate the predicted Comptonisation map using (20), which in turn gives the cluster surface brightness δI_ν through (4) and its weighted (by the primary beam) Fourier transform (10), calculated on a fine grid in the uv-plane. This is then sampled at the measured uv-postions $\boldsymbol{u}_i$. Finally, the contributions from the identified radio point sources are added to the predicted visibilities directly using (24).

The covariance matrices $\boldsymbol{C}_{\nu,\nu'}$ in (29) describe the generalised noise on the observed visibilities resulting from instrumental (receiver) noise, primordial CMB anisotropies and the background of unresolved radio point sources. Since, in this paper, we are assuming a background cosmology, the covariance matrices $\boldsymbol{C}_{\nu,\nu'}$ are not functions of the model parameters $\boldsymbol{\Theta}$ and hence need only be calculated once (similarly, the Cholesky decomposition of these matrices, required for the calculation of χ^2, need only be performed once).

Assuming the three contributions to the generalised noise are independent, the covariance matrices can be written as

$$\boldsymbol{C}_{\nu,\nu'} = \boldsymbol{C}_{\nu,\nu'}^{\mathrm{rec}} + \boldsymbol{C}_{\nu,\nu'}^{\mathrm{CMB}} + \boldsymbol{C}_{\nu,\nu'}^{\mathrm{conf}}. \tag{31}$$

The first term on the right hand side is a diagonal matrix with elements $\sigma_{\nu,i}^2\delta_{ij}\delta_{\nu\nu'}$, where $\sigma_{\nu,i}$ the rms Johnson noise on the ith element of the data vector $\boldsymbol{d}_\nu$ at frequency ν. The second term contains significant off-diagonal elements both between visibility positions and between frequencies, and can be calculated from a given primary CMB power spectrum $C_\ell^{\mathrm{CMB}}(\nu)$ following Hobson & Maisinger (2002); note that in intensity units the CMB power spectrum is a function of frequency. The third term in (31) is the covariance matrix of the source confusion noise, which allows for the remaining unresolved radio sources with flux densities less than some flux limit S_{lim} that have been left after high-resolution observation and subtraction. Scheuer (1957) was the first to show that the flux from such unknown sources could be considered statistically as a Poissonian contribution (Gaussian, in the limit of large numbers of sources; see e.g. Condon 1974) to the measured signal. The response of an interferometer to a radio source has a positive signal and both positive and negative sidelobes. Scheuer's work is easily modified to the case where all sources with fluxes greater than a limiting flux S_{lim} (perhaps from another telescope) have already been subtracted. Assuming the unresolved radio sources to

be randomly distributed on the sky leads to a flat angular power spectrum for confusion noise, given by (in intensity units)

$$C_\ell^{\mathrm{conf}}(\nu) = \int_0^{S_{\mathrm{lim}}} S^2 n_\nu(S)\, dS. \qquad (32)$$

where S_{lim} is the completeness limit (at, say, 5σ) of the source subtraction survey, and $n_\nu(S) \equiv dN_\nu(>S)/dS$ is the differential source count at frequency ν as a function of flux S. In principle, one should take into account that the unresolved radio point sources are not randomly distributed on the sky, and may be concentrated within clusters. This would lead to a power spectrum that was a function of ℓ, but we do not pursue this further here. In any case, the power spectrum $C_\ell^{\mathrm{conf}}(\nu)$ can be used to construct the corresponding confusion noise covariance matrices in the same way as for the CMB contribution (see Hobson & Maisinger 2002 for details).

In our analysis of simulated AMI data in Section 6, we assume that the differential number count $n_\nu(S)$ does not vary with frequency over the AMI band 13.9–18.2 GHz. The deepest source counts at frequencies above ≈ 5 GHz are those at 15 GHz from the 9C survey (Waldram et al. 2003), for which

$$n(S) = 51 S^{-2.15}\ \mathrm{Jy}^{-1}\,\mathrm{sr}^{-1}, \qquad (33)$$

based on 465 sources above a 5-σ completeness of 25 mJy. We use this count in our analysis.

5.4 Estimation of model parameters

The posterior $P(\boldsymbol{\Theta}) \propto L(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})$ of the model parameters $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_{\mathrm{c}}, \boldsymbol{\Psi})$ can be efficiently and robustly explored using the posterior weighted samples produced by the MULTINEST algorithm (Feroz & Hobson 2008; Feroz et al. 2008). From these samples, one can, for example, construct one-dimensional marginalised posterior distributions for each parameter, from which best-fit values and uncertainties are trivially obtained. In terms of cluster modelling, one is interested only in the cluster parameters $\boldsymbol{\Theta}_{\mathrm{c}} = (x_{\mathrm{c}}, y_{\mathrm{c}}, \beta, r_{\mathrm{c}}, T, M_{\mathrm{g}}, z)$, whereas the parameters $\boldsymbol{\Psi}$ associated with the resolved radio point sources are considered as nuisance parameters and are marginalised over. It may, however, also be of interest instead to marginalised over the cluster parameters and produce one-dimensional marginals for the flux S_0 of each resolved radio source, as well as its spectral index α.

5.5 Quantification of cluster detection

Owing primarily to the presence of primary CMB anisotropies, it is extremely important to quantify SZ cluster detection. We now discuss how one may calculate the probability that the observed field does indeed contain a real cluster above some particular mass limit of interest.

This quantification is most naturally performed via a Bayesian model selection by evaluating the evidence associated with the posterior for competing models for the data (see e.g. Hobson & McLachlan 2003). It is convenient to consider the following models (or hypotheses):

$H_0 = $ 'a cluster with $M_{\mathrm{g,min}} < M_{\mathrm{g}} \leqslant M_{\mathrm{g,lim}}$ is centred in S',

$H_1 = $ 'a cluster with $M_{\mathrm{g,lim}} < M_{\mathrm{g}} < M_{\mathrm{g,max}}$ is centred in S',

where S is the total prior region in the spatial subspace $\boldsymbol{x}_c = (x_c, y_c)$. Here $M_{\mathrm{g,min}}$ is the lower limit of our assumed prior range on the cluster gas mass; hence clusters below this minimum mass are supposed not to exist. Similarly, $M_{\mathrm{g,max}}$ is the upper limit of

our assumed prior range. Finally, $M_{\mathrm{g,lim}}$ is the limiting gas mass of interest that we discuss in more detail below.

We must calculate the model selection ratio R given in (3) between the hypotheses H_0 and H_1. For each hypothesis H_i ($i = 0, 1$), the evidence is given by

$$\mathcal{Z}_i = \int \mathcal{L}(\boldsymbol{\Theta})\pi_i(\boldsymbol{\Theta})\, d\boldsymbol{\Theta}, \qquad (34)$$

where

$$\pi_i(\boldsymbol{\Theta}) = \pi_i(\boldsymbol{x}_{\mathrm{c}})\pi_i(\beta)\pi_i(r_{\mathrm{c}})\pi_i(T)\pi_i(M_{\mathrm{g}}, z)\pi_i(\boldsymbol{\Psi}), \qquad (35)$$

for $i = 0, 1$, are priors that define the hypotheses. In particular, the priors on all the cluster parameters and source parameters, apart from M_{g} and z, may be taken to be the same as those discussed above for both hypotheses. Differences between the priors for the two hypotheses do occur in $\pi_i(M_{\mathrm{g}}, z)$, but in a straightforward manner. For hypothesis H_0, we use $\pi(M_{\mathrm{g}}, z)$ given in (22), but now appropriately normalised over the range $M_{\mathrm{g,min}} < M_{\mathrm{g}} \leqslant M_{\mathrm{g,lim}}$, and the prior is zero outside this range. Similarly, for hypothesis H_1, we use (22) appropriately normalised over the range $M_{\mathrm{g,lim}} < M_{\mathrm{g}} < M_{\mathrm{g,max}}$, and the prior is zero outside this range. The evidences (34) for $i = 0, 1$ are easily obtained using the MULTINEST algorithm.

So far we have not addressed the prior ratio $\mathrm{Pr}(H_1)/\mathrm{Pr}(H_0)$ in (3). This is easily obtained from the prior distribution $\pi(M_{\mathrm{g}}, z)$ in (22), and is given by

$$\frac{\mathrm{Pr}(H_1)}{\mathrm{Pr}(H_0)} = \frac{\int_{z_{\mathrm{min}}}^{z_{\mathrm{max}}} \int_{M_{\mathrm{g,lim}}}^{M_{\mathrm{g,max}}} \pi(M_{\mathrm{g}}, z)\, dM\, dz}{\int_{z_{\mathrm{min}}}^{z_{\mathrm{max}}} \int_{M_{\mathrm{g,min}}}^{M_{\mathrm{g,lim}}} \pi(M_{\mathrm{g}}, z)\, dM\, dz}. \qquad (36)$$

It is worth noting that, in the case where the cluster gas fraction f_{g} is assumed known, the above prior ratio is simply

$$\frac{\mathrm{Pr}(H_1)}{\mathrm{Pr}(H_0)} = \frac{\int_{z_{\mathrm{min}}}^{z_{\mathrm{max}}} \int_{M_{\mathrm{g,lim}}/f_{\mathrm{g}}}^{M_{\mathrm{g,max}}/f_{\mathrm{g}}} \frac{d^2n}{dM dz}\, dM\, dz}{\int_{z_{\mathrm{min}}}^{z_{\mathrm{max}}} \int_{M_{\mathrm{g,min}}/f_{\mathrm{g}}}^{M_{\mathrm{g,lim}}/f_{\mathrm{g}}} \frac{d^2n}{dM dz}\, dM\, dz}, \qquad (37)$$

where $d^2n/dM dz$ is the assumed cluster mass function, i.e. the distribution of the projected number density of clusters in a given mass and redshift bin per unit area. Moreover, if it is also assumed that the cluster redshift is known to be $z = z_{\mathrm{c}}$, then (37) reduces to

$$\frac{\mathrm{Pr}(H_1)}{\mathrm{Pr}(H_0)} = \frac{\int_{M_{\mathrm{g,lim}}/f_{\mathrm{g}}}^{M_{\mathrm{g,max}}/f_{\mathrm{g}}} \frac{dn}{dM}\big|_{z=z_{\mathrm{c}}}\, dM}{\int_{M_{\mathrm{g,min}}/f_{\mathrm{g}}}^{M_{\mathrm{g,lim}}/f_{\mathrm{g}}} \frac{dn}{dM}\big|_{z=z_{\mathrm{c}}}\, dM}. \qquad (38)$$

We are thus able to calculate the model selection ratio R in (3), which gives us the relative probability that the field contains a 'true' cluster, with gas mass above the limit $M_{\mathrm{g,lim}}$, as opposed to 'false' cluster, with gas mass below this limit. This, in turn, allows us to calculate the probability that the field contains a 'true' cluster, which is given by

$$p = \frac{R}{1 + R}. \qquad (39)$$

6 APPLICATION TO SIMULATED SZ OBSERVATION

In this section we describe the results of our SZ cluster modelling algorithm on simulated SZ cluster data-sets from the Arcminute Microkelvin Imager (AMI).
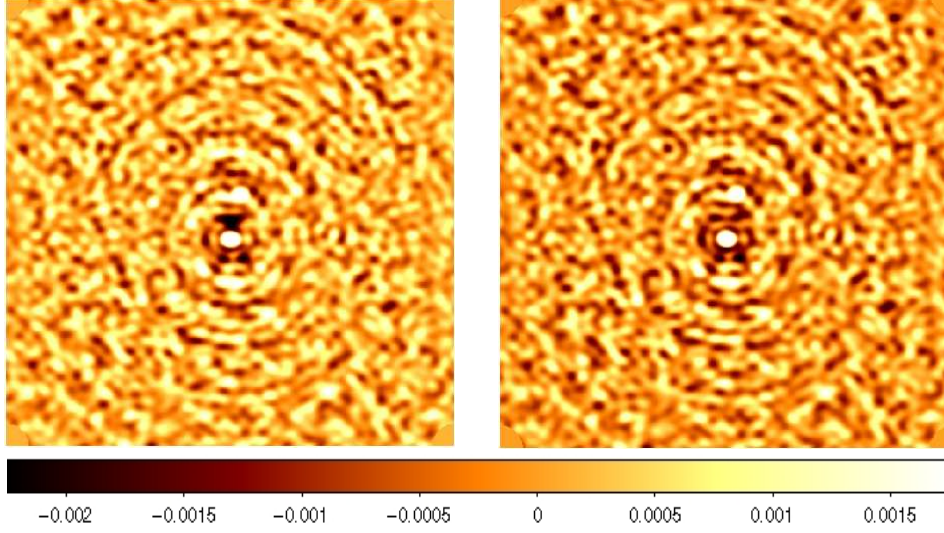
Figure 3. Maps made from channel 4 (14.992 GHz) of the simulations considered in this work with Simulation 'A' in the left panel and Simulation 'B' in the right panel. The unCLEANed maps are 512×512 pixels at $15''$ resolution. Both simulations have the same realisation of primary CMB and instrumental noise. The three radio sources are also the same between the two simulations. Simulation 'A' has a spherically-symmetric, isothermal β-model cluster at the centre of the map. Simulation 'B' has no cluster.

6.1 Simulated AMI data-sets

In simulating mock skies and observing them with a model AMI SA, we have used the methods outlined in Hobson & Maisinger (2002) and Grainge et al. (2002). We consider two simulations. Simulation 'A' has a cluster at $z = 0.3$ modelled as a spherically-symmetric isothermal β-profile with $r_c = 60''$, $\beta = 0.65$, $n_e = 10^{-2}$ cm^{-3} and $T = 8$ keV. The gas profile is linearly tapered to zero between $20r_c$ and $20.01r_c$. The Comptonisation y-parameter for this model is evaluated on a cube whose face has 512×512 pixels at $30''$ resolution before being integrated along the line of sight. Radio point sources are added to these maps using the fluxes, positions and spectral indices given in Table 2.

The uv-positions of visibility points are simulated by calculating the baselines (assuming SA antenna positions in AMI Consortium: Zwart et al. 2008) for a target at right ascension $\alpha = 4$ hours and declination $\delta = +40^o$ observed over hour angle ± 4 hours with one-second sampling.

For each simulation, a realisation of the primary CMB is calculated using a power spectrum of primary anisotropies was generated for $\ell < 8000$ using CAMB (Lewis et al. 2000), with a ΛCDM cosmology ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\sigma_8 = 0.8$ and $h = 0.7$) assumed. Primary CMB modes on ℓ scales outside the range measurable ($\ell \approx 500$–8000, considering the most extreme frequency channels) by the SA are set to zero. The CMB realisation is co-added to the cluster and radio source map in brightness temperature. To each model sky we also add a population of faint, confus-

ing radio point sources, uniformly distributed on the sky but drawn from a Poisson distribution in flux, with the 9C source count (see section 5.3) between $10\,\mu$Jy and $S_{lim} = 200\,\mu$Jy. The map is scaled by the primary beam appropriate to the measured value in that frequency channel and transformed into the Fourier plane (equivalent to Fourier transforming and convolving with the aperture illumination function). The resulting function is sampled at the required visibility points and thermal receiver noise, appropriate to the measured sensitivity of the SA, is added at this stage.

The whole process above is repeated for each of the six frequency channels. Simulation 'B' is identical to Simulation 'A', but has no cluster. Maps made from the simulated visibilities for channel 4 (14.992 GHz), for both models, are shown in Figure 3.

6.2 Analysis and results

We analysed the cluster simulations discussed above assuming a cluster model with spherical geometry, a beta profile for the gas and isothermal temperature. The priors used are listed for convenience in Table 3. Positions of the radio point sources were fixed to their true values. The primordial CMB and confusion noise were included through the covariance matrix as discussed in Section 5.3. For the confusion noise we used the 9C source count with limiting flux $S_{lim} = 200\mu$Jy (correctly) and the count cut off below $10\,\mu$Jy.

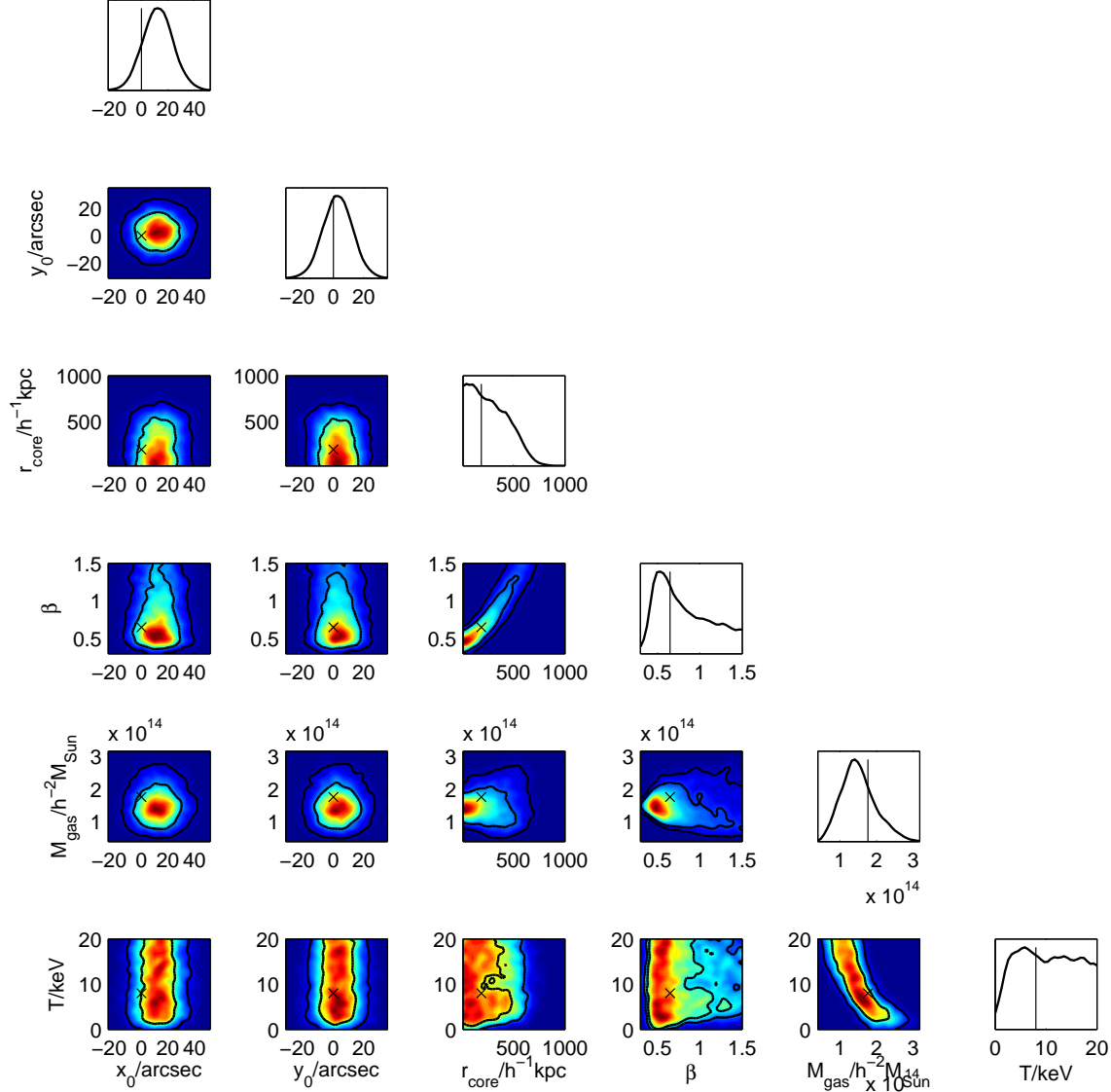| Parameters	Priors
x_c, y_c	$(0 \pm 60)''$
M_g	$12 < \log_{10} M_g/h^{-2}M_\odot < 14.5$
T	$0 < T/\text{keV} < 20$
r_c	$0 < r_c/h^{-1}\,\text{kpc} < 1000$
β	$0.3 < \beta < 1.5$
S_0	$0 < S/\text{mJy} < 20$

Table 3. Priors for the cluster and source parameters. Inequalities denote uniform prior probability between the given limits, whilst $(a \pm b)$ denotes a Gaussian prior with mean a and variance b^2.

	Δx/arcsec	Δy/arcsec	S_{15}/mJy	α
1	8	10	5	0
2	0	–5	15	+1 (falling)
3	–3	8	8	–0.3 (rising)

Table 2. Contaminating radio sources for the simulations considered in this work. Source positions are given in arcminute offsets from the pointing centre. The flux and spectral index are at 15.0 GHz.

Figure 4. 2-D marginalized posterior probability distributions for the parameters of the cluster Simulation 'A' discussed in Section 6. The true parameter values used in the simulation are shown by crosses and vertical lines in 2-D and 1-D marginalisations respectively.

For the primordial CMB anisotropies, we (correctly) assumed a ΛCDM cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\sigma_8 = 0.8$ and $h = 0.7$. We also assumed that the redshift of the cluster $z = 0.3$ and the gas fraction $f_g = 0.1$ were known. We analysed all six AMI frequencies channels jointly. In quantifying our cluster detection, we adopted a gas mass limit of $M_{g,\text{lim}} = 10^{13}h^{-1}M_\odot$ and assumed a Press–Schechter mass function.

For the Simulation 'B' (with radio point sources and no cluster) MULTINEST did identify a dominant peak in the posterior distribution of cluster parameters but the probability odds ratio R, as discussed in Section 5.5, was evaluated to be 0.32 ± 0.03, showing that it is more than twice as likely that the field did not contain a 'true' cluster with a gas mass above the mass limit of interest. Since there is no cluster in the field, the highest likelihood point comes from a large negative primordial feature, but since the statistics of the primordial CMB have been incorporated in the likelihood evaluation through the covariance matrix, the Bayesian model selection takes this into account and consequently the odds ratio is in favour

of the detected feature being 'false'. To verify this assertion, we analysed Simulation 'B' without including the CMB component in the covariance matrix, in which case the probability odds ratio for cluster detection R was evaluated to be ≈ 150 which clearly shows that including the CMB is extremely important for properly modelling galaxy clusters through the SZ effect.

For the Simulation 'A' (with cluster and radio point sources), the probability odds ratio for cluster detection R was evaluated to be $e^{12.2 \pm 0.2}$, showing an overwhelming evidence in favour of a 'true' cluster detection. We plot the 2-D and 1-D marginalized posterior distributions of the cluster parameters along with the true parameter values used in the simulation in Figures 4 and 5 respectively. In Figure 5, we also plot the prior distributions imposed on the parameters. The inferred cluster parameter means and 1-σ uncertainties are listed in Table 4. It is clear from this table and the posterior plots that all the model parameters have been estimated to reasonable accuracy. It can also be seen that the posterior for the cluster temperature is largely unconstrained and the expected
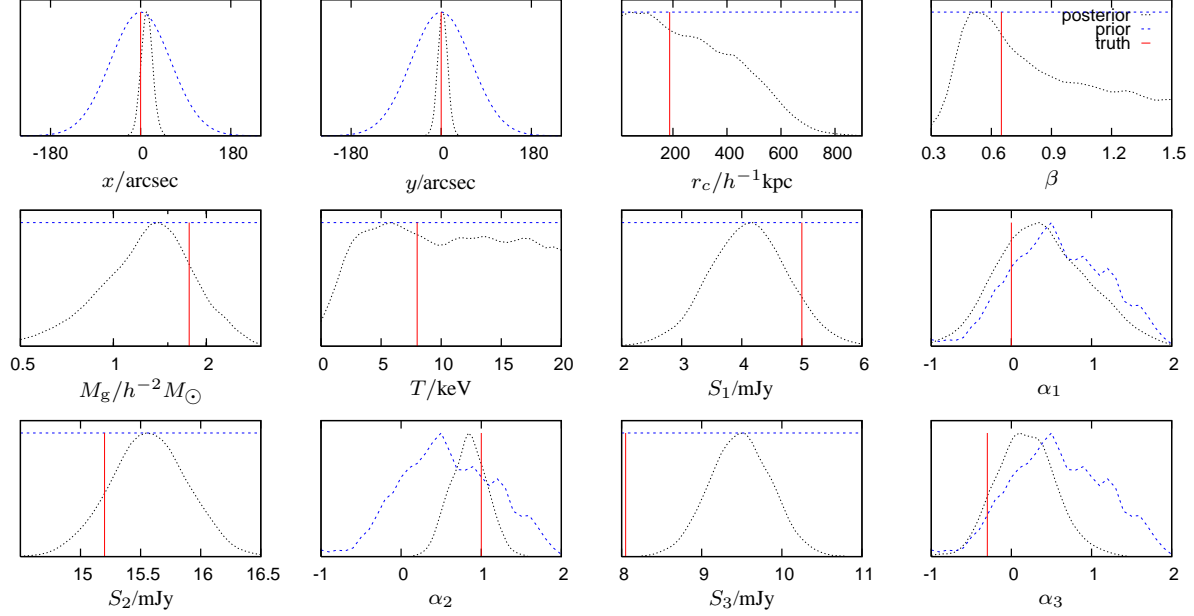
Figure 5. Priors and 1-D marginalized posterior probability distributions for the parameters of the simulated cluster and radio point sources discussed in Section 6. The true parameter values used in the simulation are shown by vertical lines.

degeneracy between T and M_g is also evident. Clearly some additional information on cluster T is required to get a sensible estimate for M_g. This information can come from the X-ray observation of the same cluster or from an optical velocity dispersion measurement. T calculated in such a way can be used as a prior for the analysis of the cluster. The result of a temperature measurement of (8 ± 2) keV applied as a Gaussian prior on T is shown in Figure 6 and 7.

It is also of interest to investigate the importance of modelling the three radio point sources, with properties given in Table 2, that were added to Simulation 'A'. Assuming known source positions and allowing the source fluxes to vary, results in a massive increase in evidence over that for an analysis (incorrectly) assuming no radio point sources.

$$\frac{P(\boldsymbol{D}|3\,\text{sources})}{P(\boldsymbol{D}|\text{No sources})} = e^{961.9 \pm 0.2}. \qquad (40)$$

We list the inferred cluster parameters for the analysis (incorrectly) assuming no radio point sources in Table 5. Comparing the parameter values in Table 5 with the values used in the simulation, it is

Parameters	Inferred values
x_c/arcsec	12 ± 11
y_c/arcsec	3 ± 10
r_c/h^{-1}kpc	276 ± 177
β	0.8 ± 0.3
$M_\mathrm{g}/h^{-2}M_\odot$	$(1.5 \pm 0.5) \times 10^{14}$
T/keV	10 ± 5

Table 4. Inferred cluster parameters values for the analysis with cluster Simulation 'A' discussed in Section 6. All noise sources discussed in Section 5.3 were included in the covariance matrix and the observed radio sources were treated as nuisance parameters. The probability odds ratio for cluster detection R was evaluated to be $e^{12.2 \pm 0.2}$.
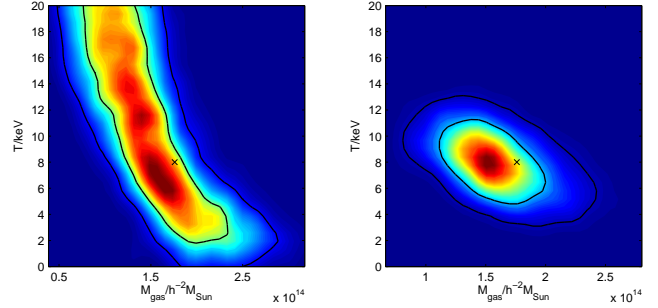
Figure 6. 2-D marginalized posterior probability distributions for T and M_g for the analysis of the cluster Simulation 'A' discussed in Section 6. The true parameter values used in the simulation are shown by crosses. Uniform $\mathcal{U}(0, 20)$ keV and Gaussian (8 ± 2) keV priors on T were used for the figures on the left and right hand-panels respectively.
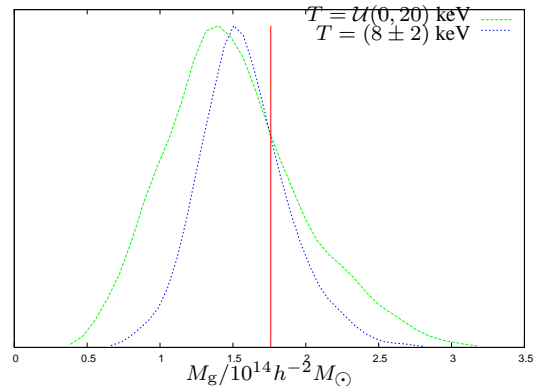
Figure 7. 1-D marginalized posterior probability distributions for M_g for the analysis of the cluster Simulation 'A' discussed in Section 6. The true M_g used in the simulation is shown by the vertical line.

clear that the model used in the analysis is incorrect. The algorithm is trying to fit for the cluster assuming no radio point sources in the field, but the presence of radio point sources is forcing the algorithm to compensate by preferring larger cluster r_c and β. This highlights the importance of radio point souce information while analyzing the SZ data.

Finally, to investigate the importance of including the noise contribution due to unsubtracted radio sources, we analysed Simulation 'A' again but this time without including the contribution from the confusion noise to the covariance matrix. The priors on the cluster and radio source parameters used were the same as listed in Table 3. This resulted in the probability odds ratio for cluster detection R of $e^{50.8\pm0.2}$, again showing a very strong evidence in favour of a true cluster detection. We list the inferred cluster parameters for this analysis in Table 6. It can be seen from this table that while most of the cluster parameters have been inferred to reasonable accuracy, the inferred cluster position is 5σ away from the true centre.

Parameters	Inferred values
x_c/arcsec	13 ± 10
y_c/arcsec	40 ± 8
r_c/h^{-1}kpc	296 ± 175
β	0.8 ± 0.3
$M_{\rm g}/h^{-2}M_\odot$	$(1.6 \pm 0.5) \times 10^{14}$
T/keV	10 ± 5

Table 6. Inferred cluster parameters values for the analysis with cluster simulation 'A' discussed in Section 6 and (incorrectly) assuming no unsubtracted radio sources below $S_{\rm lim}$.

there are several degeneracies in cluster parameters when modelled through the SZ effect only, the most significant of which are between r_c and β, and between T and $M_{\rm g}$. Better constraints on these parameters can be obtained by using information from different waveband observations.

7 CONCLUSIONS

Extracting and parametrizing clusters in SZ data is an extremely challenging task due to the presence of (both observed and unsubtracted) contaminating radio sources and primary CMB anisotropies. We have described an efficient approach using the MULTINEST algorithm to model galaxy clusters in multi-frequency pointed SZ observations, in the presence of radio point sources. The parameters of the radio sources are treated as nuisance parameters, which allows for the fitting of these parameters simultaneously with the cluster parameters and consequently for the uncertainties in the measurements of the radio source parameters to be propagated to the cluster parameter inferences. We considered the three main sources of noise for SZ observations: (a) receiver noise, (b) primary CMB anisotropies and (c) confusion noise due to unsubtracted radio sources. We have also shown that it is extremely important to take into account all these noise contributions in the analysis to get the correct posterior distributions for the cluster parameters. Even with this extra complexity, we are able to analyse a pointed SZ cluster observation with six frequency channels on a single Intel Woodcrest 3.0-GHz processor in about five hours. The work presented in this paper is limited to pointed cluster observations. We plan to extend this methodology to multi-field survey observations in a future study. The code is fully parallel, making our algorithm a viable option for even the deepest SZ surveys.

Our analysis methodology is easily extendable to do a joint analysis of clusters using SZ, lensing and X-ray data-sets. This data fusion is extremely important for understanding cluster physics, as

Parameters	Inferred values
x_c/arcsec	14 ± 10
y_c/arcsec	24 ± 8
r_c/h^{-1}kpc	546 ± 65
β	1.4 ± 0.1
$M_{\rm g}/h^{-2}M_\odot$	$(2.1 \pm 0.8) \times 10^{14}$
T/keV	11 ± 5

Table 5. Inferred cluster parameters values for the analysis with cluster Simulation 'A' discussed in Section 6 and (incorrectly) assuming no observed radio point sources.

REFERENCES

AMI Consortium: Grainge K., et al., 2006, MNRAS, 369, L1
AMI Consortium: Zwart J. T. L., et al., 2008, MNRAS accepted, preprint (astro-ph/0807.2469), 807
Battye R. A., Weller J., 2003, Phys.Rev.D, 68, 083506
Birkinshaw M., 1999, Phys.Rep., 310, 97
Birkinshaw M., Gull S. F., Hardebeck H., 1984, Nature, 309, 34
Birkinshaw M., Gull S. F., Moffet A. T., 1981, ApJ, 251, L69
Bonamente M., Joy M. K., LaRoque S. J., Carlstrom J. E., Reese E. D., Dawson K. S., 2006, ApJ, 647, 25
Carlstrom J. E., Holder G. P., Reese E. D., 2002, ARA&A, 40, 643
Carlstrom J. E., Joy M., Grego L., 1996, ApJ, 456, L75
Cavaliere A., Fusco-Femiano R., 1976, A&A, 49, 137
Cavaliere A., Fusco-Femiano R., 1978, A&A, 70, 677
Condon J. J., 1974, ApJ, 188, 279
Cotter G., Buttery H. J., Das R., Jones M. E., Grainge K., Pooley G. G., Saunders R., 2002, MNRAS, 334, 323
Cotter G., et al., 2002, MNRAS, 331, 1
Dobbs M., Halverson N. W., Ade P. A. R., Basu K., Beelen A., Bertoldi F., Cohalan C., 2006, New Astronomy Review, 50, 960
Evrard A. E., et al., 2002, ApJ, 573, 7
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449

Feroz F., Hobson M. P., Bridges M., 2008, preprint (astro-ph/0807.2469)

Grainge K., Grainger W. F., Jones M. E., Kneissl R., Pooley G. G., Saunders R., 2002, MNRAS, 329, 890

Grainge K., Jones M., Pooley G., Saunders R., Baker J., Haynes T., Edge A., 1996, MNRAS, 278, L17

Grainge K., Jones M. E., Pooley G., Saunders R., Edge A., Grainger W. F., Kneissl R., 2002, MNRAS, 333, 318

Grainger W. F., Das R., Grainge K., Jones M. E., Kneissl R., Pooley G. G., Saunders R. D. E., 2002, MNRAS, 337, 1207

Grego L., Carlstrom J. E., Joy M. K., Reese E. D., Holder G. P., Patel S., Cooray A. R., Holzapfel W. L., 2000, ApJ, 539, 39

Hobson M. P., Bridle S. L., Lahav O., 2002, MNRAS, 335, 377

Hobson M. P., Maisinger K., 2002, MNRAS, 334, 569

Hobson M. P., McLachlan C., 2003, MNRAS, 338, 765

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372

Jones M. E., et al., 2005, MNRAS, 357, 518

Jones M. E., Saunders R., Alexander P., Birkinshaw M., Dillon N., Grainge K., Hancock S., Lasenby A., Lefebvre D., Pooley G., 1993, Nature, 365, 320

Joy M., et al., 2001, ApJ, 551, L1

Kneissl R., Jones M. E., Saunders R., Eke V. R., Lasenby A. N., Grainge K., Cotter G., 2001, MNRAS, 328, 783

Komatsu E., Seljak U., 2001, MNRAS, 327, 1353

Kosowsky A., 2006, New Astronomy Review, 50, 969

LaRoque S. J., Bonamente M., Carlstrom J. E., Joy M. K., Nagai D., Reese E. D., Dawson K. S., 2006, ApJ, 652, 917

Lay O. P., Halverson N. W., 2000, ApJ, 543, 787

Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473

Li C.-T., et al., 2006, in Millimeter and Submillimeter Detectors and Instrumentation for Astronomy III. Edited by Zmuidzinas, Jonas; Holland, Wayne S.; Withington, Stafford; Duncan, William D. Vol. 6275 of Proceedings of the SPIE

Liddle A. R., 2007, MNRAS, 377, L74

Marshall P. J., 2003, PhD thesis, University of Cambridge

Marshall P. J., Hobson M. P., Slosar A., 2003, MNRAS, 346, 489

Mason B. S., Myers S. T., 2000, ApJ, 540, 614

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Patel S. K., et al., 2000, ApJ, 541, 37

Press W. H., Schechter P., 1974, ApJ, 187, 425

Reese E. D., Carlstrom J. E., Joy M., Mohr J. J., Grego L., Holzapfel W. L., 2002, ApJ, 581, 53

Reese E. D., et al., 2000, ApJ, 533, 38

Ruhl J., et al., 2004, in Bradford C. M., Ade P. A. R., Aguirre J. E., Bock J. J., Dragovan M., Duband L., Earle L., Glenn J., Matsuhara H., Naylor B. J., Nguyen H. T., Yun M., Zmuidzinas J., eds, Millimeter and Submillimeter Detectors for Astronomy II. Edited by Jonas Zmuidzinas, Wayne S. Holland and Stafford Withington Vol. 5498 of Proceedings of the SPIE. p. 11

Sarazin C., 1988, X-ray Emission from Clusters of Galaxies. Cambridge: CUP

Saunders R., et al., 2003, MNRAS, 341, 937

Scheuer P. A. G., 1957, Proc. Cambridge Phil. Soc., 53, 764

Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1

Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, American Institute of Physics Conference Series Nested Sampling. pp 395–405

Sunyaev R. A., Zeldovich Y. B., 1970, CoASP, 2, 66

Sunyaev R. A., Zeldovich Y. B., 1972, CoASP, 4, 173

Trotta R., 2007, MNRAS, 378, 72

Uson J. M., 1986, in O'Dea C. P., Uson J. M., eds, Radio Continuum Processes in Clusters of Galaxies The Sunyaev-Zel'dovich effect – Measurements and implications. p. 255

Waldram E. M., Bolton R. C., Pooley G. G., Riley J. M., 2007, MNRAS, 379, 1442

Waldram E. M., Pooley G. G., Grainge K. J. B., Jones M. E., Saunders R. D. E., Scott P. F., Taylor A. C., 2003, MNRAS, 342, 915

Watson R. A., et al., 2003, MNRAS, 341, 1057

White S. D. M., Navarro J. F., Evrard A. E., Frenk C. S., 1993, Nature, 366, 429

This paper has been typeset from a TeX/ LaTeX file prepared by the author.